

# **TAU IDENTIFICATION USING A BDT-BASED ALGORITHM**

## **Preliminary studies**

**CDF-Italy  
Pierluigi Totaro  
Fermilab, 11-05-08**





# SUMMARY

- **Some definitions (datasets, matching, tau-IDs)**
- **Description of the procedure**
- **Results**



## DATASETS USED FOR THIS STUDY:

MC ( $W \rightarrow \tau \nu$ ):	we0sat	#events: 6931813
QCD (JET_20):	gjt1ah	#events: 12208412

## EVENT SELECTION:

- JET\_20 trigger emulation;
- “Unreconstructed  $\text{Pi}^0$ s” tau correction (note 6654) applied both to MC and data;
- Hadronic energy of taus corrected in MC (note 8809).

## MATCHING:

- **MC:** generated hadronic taus matched with reconstructed taus within  $\text{DR} < 0.2$
- **DATA:** jets (cone 0.4) matched with reconstructed taus within  $\text{DR} < 0.4$

# STANDARD TAU RECONSTRUCTION: two-cone algorithm for tracks and $\pi^0$ 's

## •TauFinder Level:

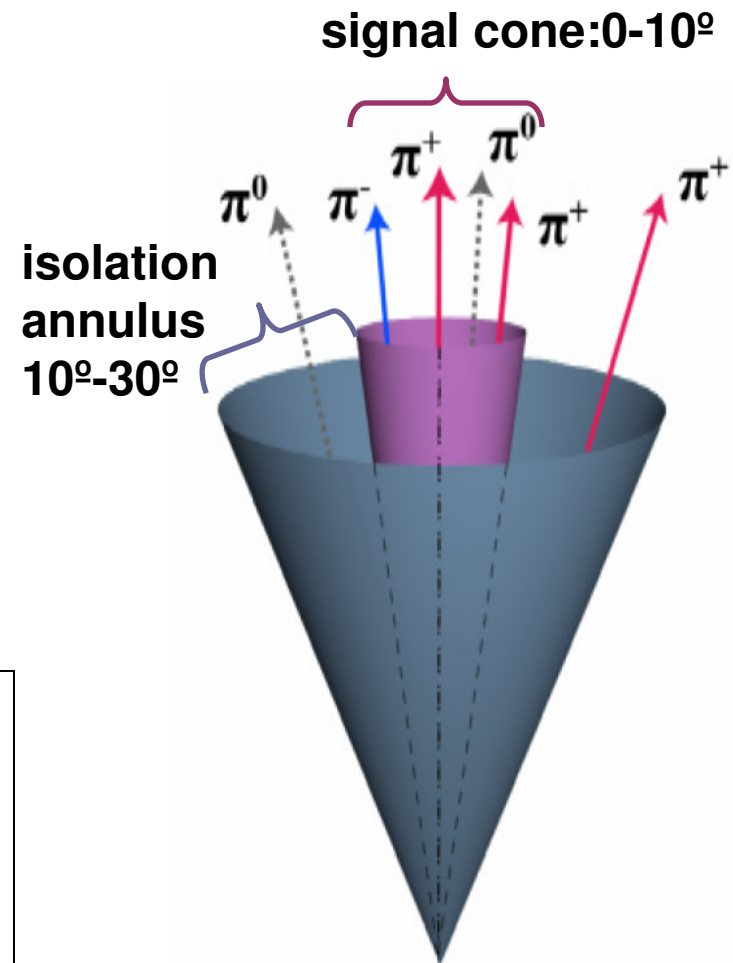
$$\begin{aligned} E_{T}^{\text{seedtrk}} &> 4.0 \text{ GeV} \\ E_{T}^{\text{sh}} &> 1.0 \text{ GeV} \\ N_{\text{trk}} &\leq 6 \\ P_{T}^{\text{seedtrk}} &> 4.5 \text{ GeV/c} \\ P_{T}^{\text{shtrk}} &> 1.0 \text{ GeV/c} \\ D_z^{\text{shtrk}} &< 5 \text{ cm} \\ |\eta| &< 1.3 \end{aligned}$$

## •Tau Selection (one of them)

see note 8639, MSSM  $H \rightarrow \tau\tau$

$$\begin{aligned} 9.0 &< |Z_{\text{ces}}| < 230.0 \text{ cm} \\ E_{T}^{\text{clus}} &> 9.0 \text{ GeV} \\ \text{Visible } E_T &> 15 \text{ (20) GeV for 1(3) prong} \\ |Q^{\text{trk}}| &= 1 \\ N_{\text{sig}}^{\text{trk}} &= 1, 3 \\ \left. \begin{aligned} P_{T}^{\text{trk}} &< 2.0 \text{ GeV/c} \\ P_{T}^{\text{Pi0}} &< 1 \text{ GeV/c} \end{aligned} \right\} &\text{isolation requirements} \\ \text{Visible Mass} &< 1.8 \text{ (2.2) for 1(3) prong} \end{aligned}$$

$$\text{Electron veto: } \xi = \frac{E_{\text{tot}}}{\sum |\vec{p}|} \left( 0.95 - \frac{E_{\text{em}}}{E_{\text{tot}}} \right) > 0.1$$

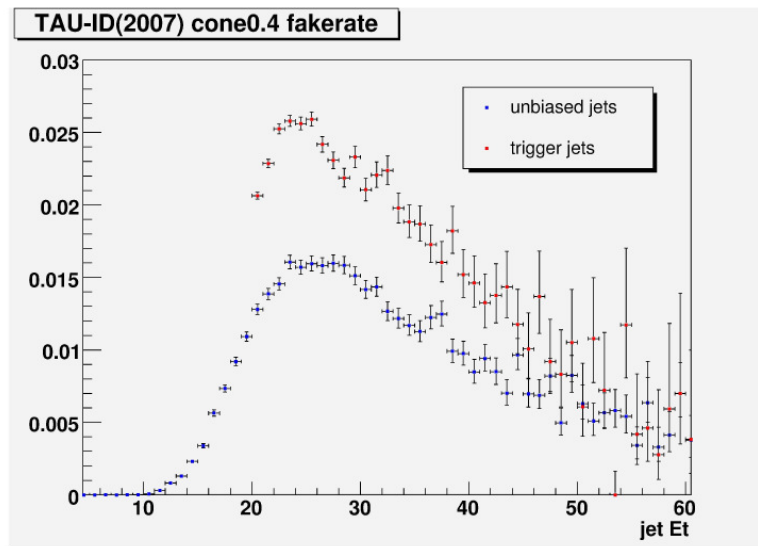


I have chosen a constant signal cone size;  
it can also be tau's energy dependent;

Slide 4

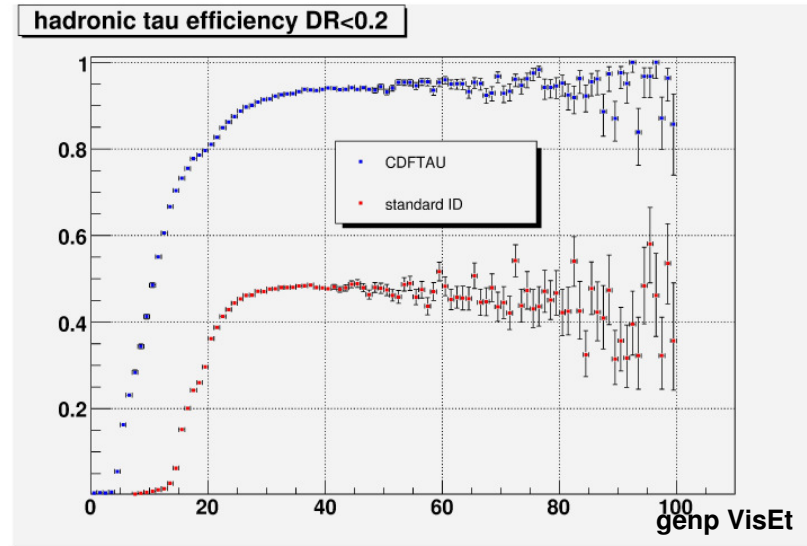
## Fake rate and efficiency using standard TAU ID definition

$$f^{\tau} = \frac{N_{rec}^{\tau}}{N^{jet} (|\eta_{det}^{jet}| < 1.0)}$$



**trigger jets:** they satisfy all 3 trigger level requirements  
**unbiased jets:** not essential for triggering the event

$$\varepsilon^{\tau} = \frac{N_{rec}^{\tau}}{N_{gen}^{\tau} (|\eta^{\tau}| < 1.0)}$$



Overall identification efficiency at plateau is about 50%



# BDT STRATEGY FOR TAU-ID IMPROVEMENT

PRELIMINARY: we define a very loose tau-ID, considering 1-prong and 3-prong cases separately

GOAL: we want to distinguish **between MC taus and tau-looking jets (fakes)**

CRITERIA: We require the jet->tau fake rate to be the same as standard ID algorithm, and we see if MC-tau efficiency can be increased, at the same time

CAUTIONS: We paid attention to remove all possible biases given by the trigger used for this study

- 1) Only un-biased jets in data sample have been considered for tau matching
- 2) We selected events with flat Visible Et distribution
- 3) We divided the energy spectrum in several subranges and treated them separately

## MY “LOOSE” TAU-ID DEFINITION:

### Tau-Finder level

$E_{T}^{\text{seedtrk}} > 4.0 \text{ GeV}$   
 $E_{T}^{\text{sh}} > 1.0 \text{ GeV}$   
 $N_{\text{twr}} \leq 6$   
 $P_{T}^{\text{seedtrk}} > 4.5 \text{ GeV}/c$   
 $P_{T}^{\text{shtrk}} > 1.0 \text{ GeV}/c$   
 $D_z^{\text{shtrk}} < 5 \text{ cm}$   
 $|\eta| < 1.3$

+

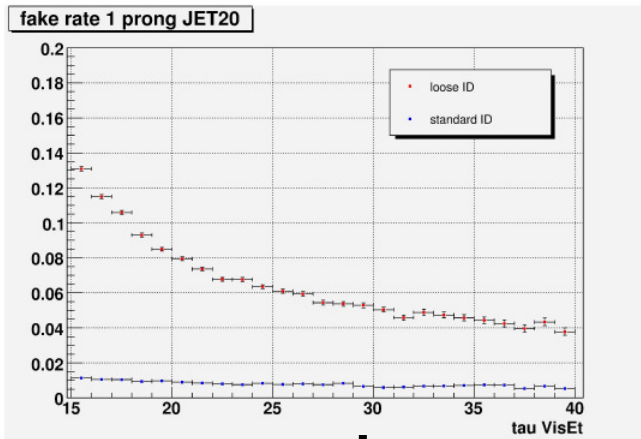
$E_{T}^{\text{seedtrk}} > 4.5 \text{ GeV}$   
 $N_{\text{sig}}^{\text{trk}} = 1, 3$   
 $|Q_{\text{trk}}| = 1$   
 $\text{Visible Mass} < 5 \text{ GeV}$   
 $|\eta| < 1.0$   
 $9 < |Z_{\text{Ces}}| < 216$   
 $\text{Electron veto: } \xi = \frac{E_{\text{tot}}}{\sum |\vec{p}|} \left(1 - \frac{E_{\text{em}}}{E_{\text{tot}}}\right) > 0.1$

### TAU VARIABLES USED IN BDT:

We consider only quantities which are considered well-modelled by MC (see Pasha's note 8809): NO Hadronic Energy, NO isolation energy

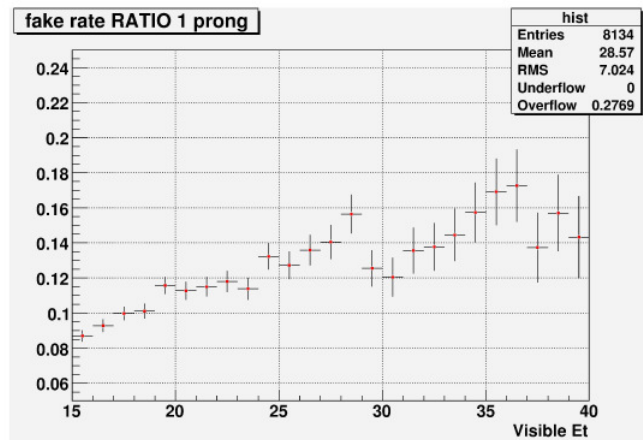
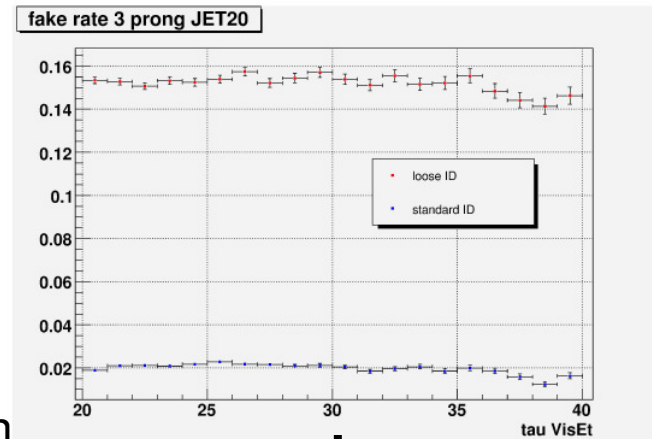
• NTracks1030, NPi01030, NPi010, SumPt30, SumPt10, SumPtPi01030, Visible Mass, EtEm, Visible Et

LET'S QUANTIFY OUR BENCHMARK:  
we want the **same fake rate as standard ID**, so

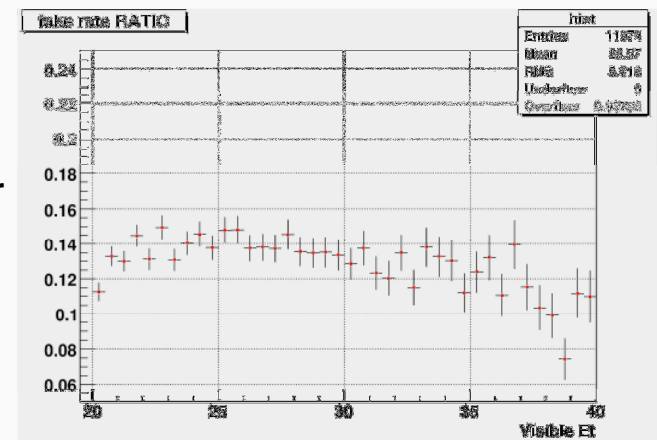


from 1-prong and 3-prong fake rates  
(loose tau and std tau)

if we divide bin by bin

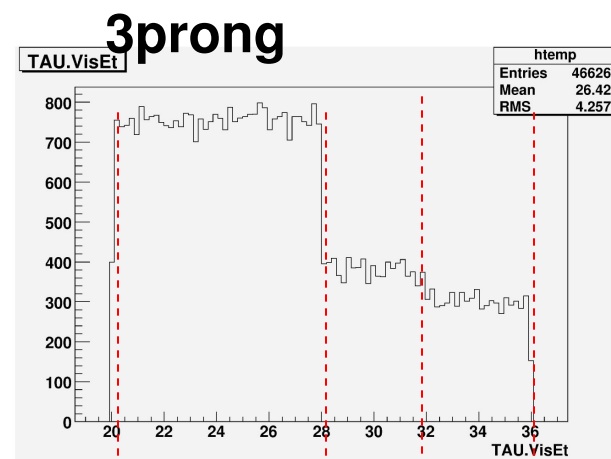
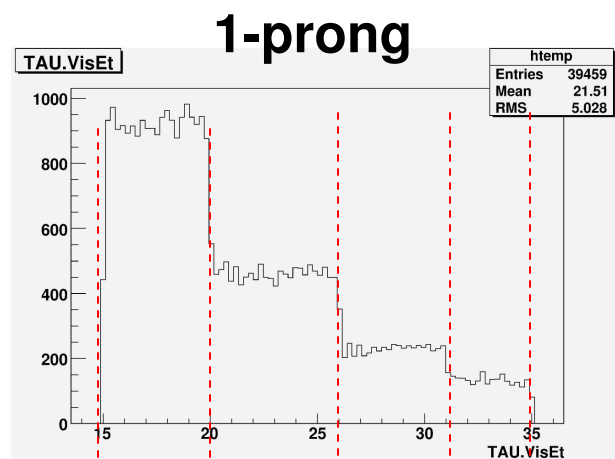


we have the bkg  
reduction factor we  
should impose in our  
BDT

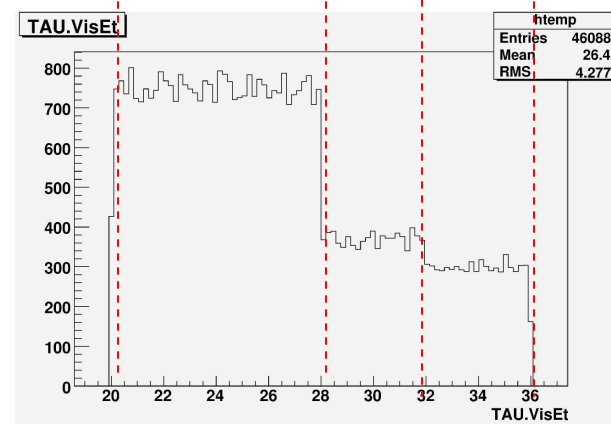
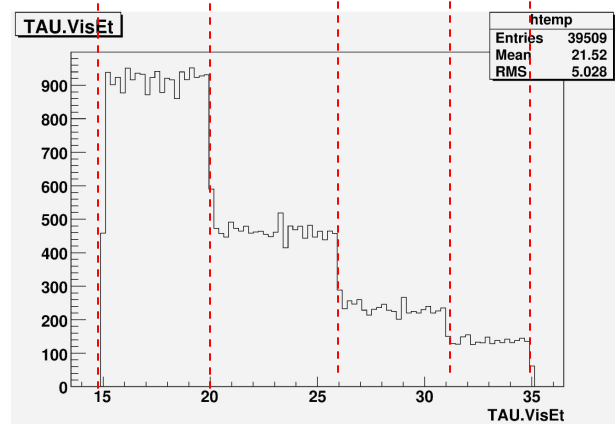


# Samples for BDT training

MC



DATA



- 15-20 GeV: 20000 ev.
- 20-26 GeV: 12000 ev.
- 26-31 GeV: 5000 ev.
- 31-35 GeV: 2400 ev.

- 20-28 GeV: 32000 ev.
- 28-32 GeV: 8000 ev.
- 32-36 GeV: 6400 ev.

Slide 9

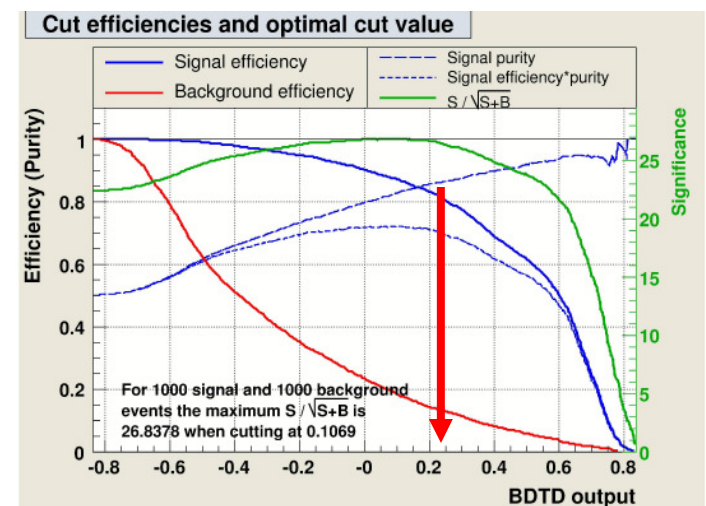
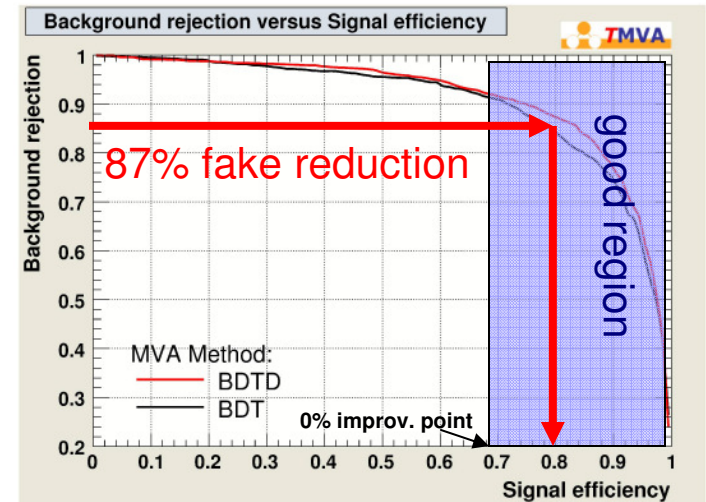
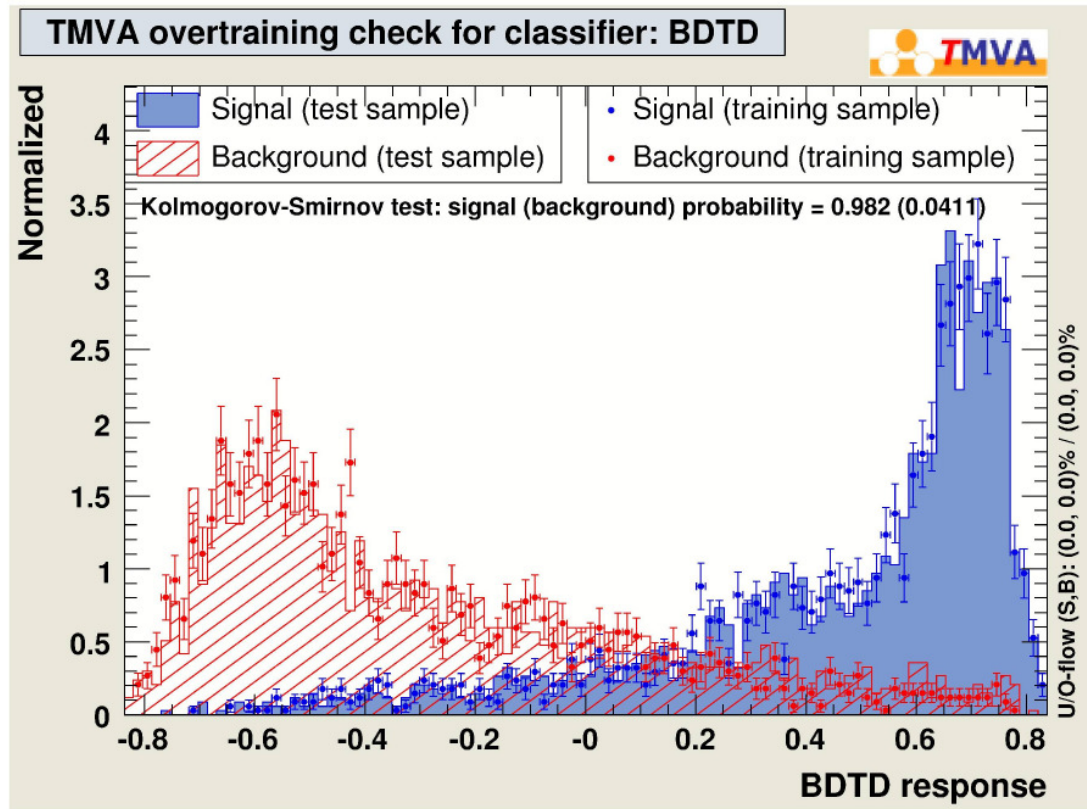
# BDT RESULTS

1 PRONG Subrange 24-26 GeV in Visible Et

Samples for training: MC #ev. ~4000 (500/250 MeV)

data #ev. ~4000 (500/250 MeV)

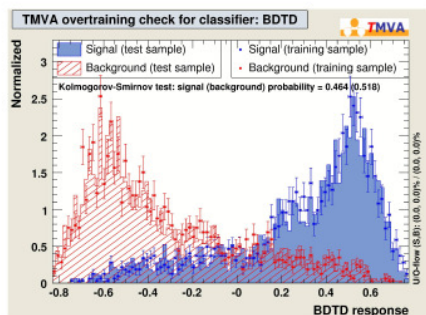
Randomly splitted into two subsamples  
for internal BDT overtraining test



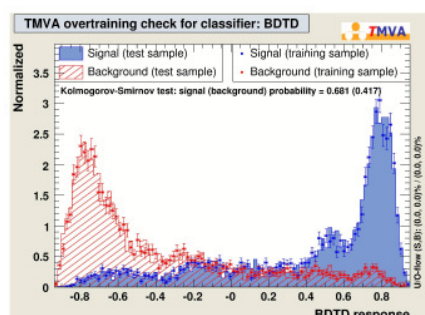
BDT with decorrelation of variables  
(BDTD) works better.

Slide 10

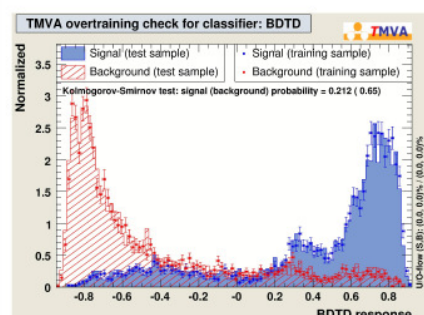
# The same procedure is applied to other VisibleEt subranges 1-prong



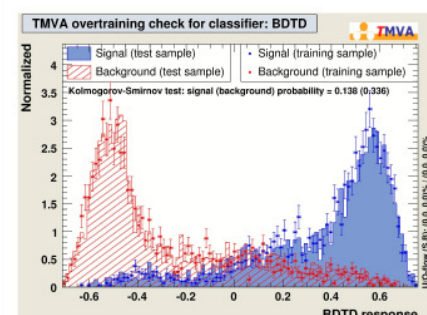
15-16 GeV  
# ev. 4000



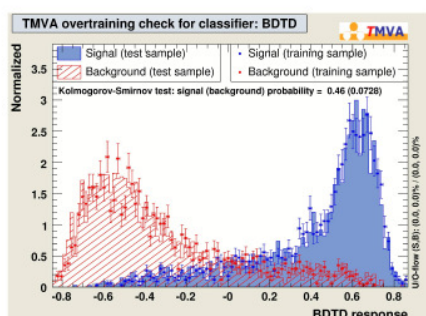
16 – 18 GeV  
# ev. 8000



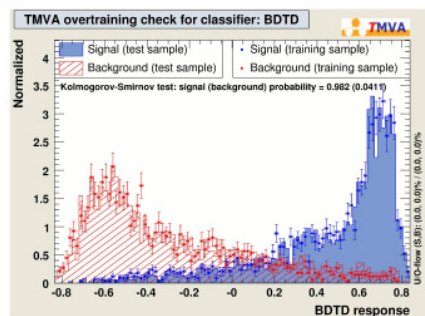
18 – 20 GeV  
# ev. 8000



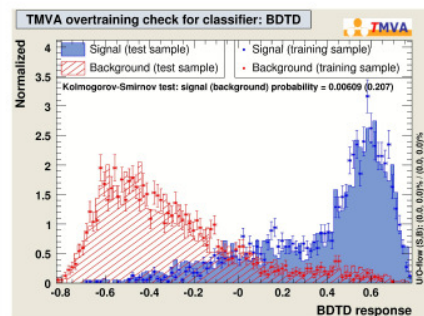
20 – 22 GeV  
# ev. 4000



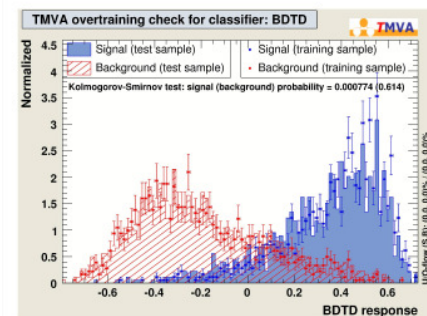
22 – 24 GeV  
# ev. 4000



24 – 26 GeV  
# ev. 4000



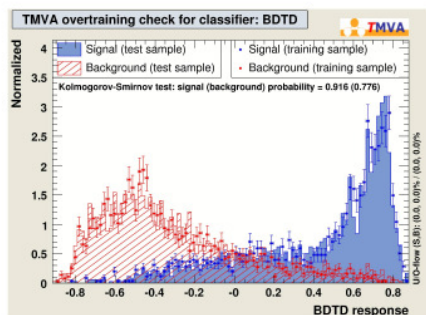
26 – 31 GeV  
# ev. 5000



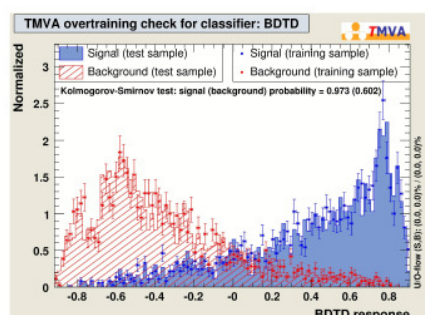
31 – 35 GeV  
# ev. 2400

Slide 11

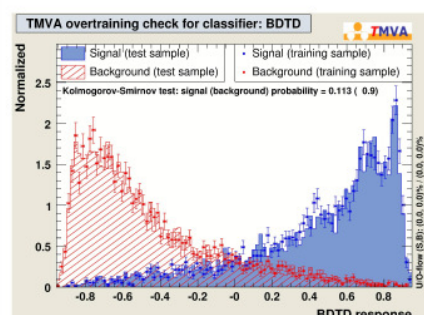
# ...and also for 3-prong case



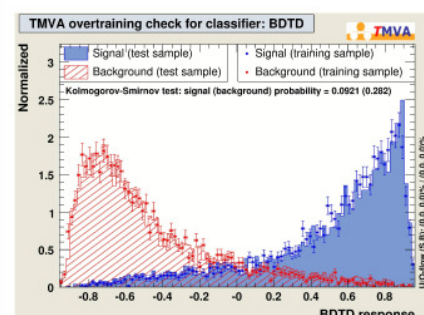
20-21 GeV  
# ev. 4000



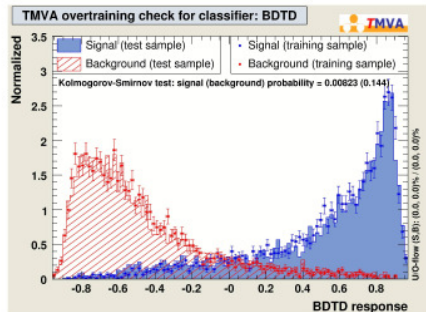
21 – 22 GeV  
# ev. 4000



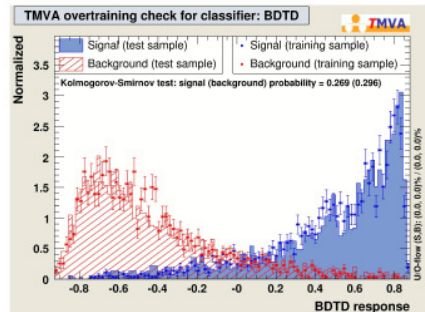
22 – 24 GeV  
# ev. 8000



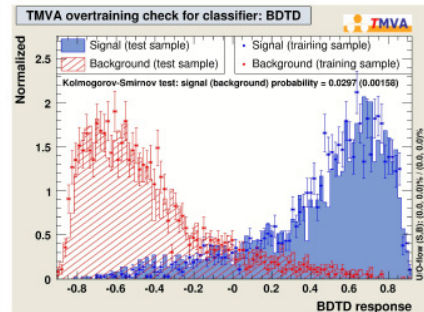
24 – 26 GeV  
# ev. 4000



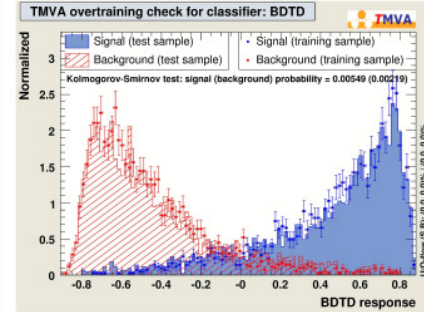
26 – 28 GeV  
# ev. 8000



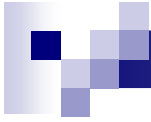
28 – 30 GeV  
# ev. 4000



30 – 32 GeV  
# ev. 4000

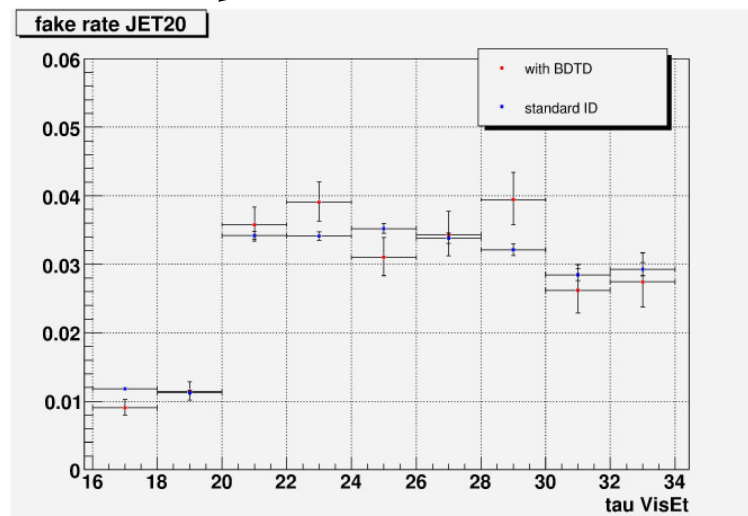
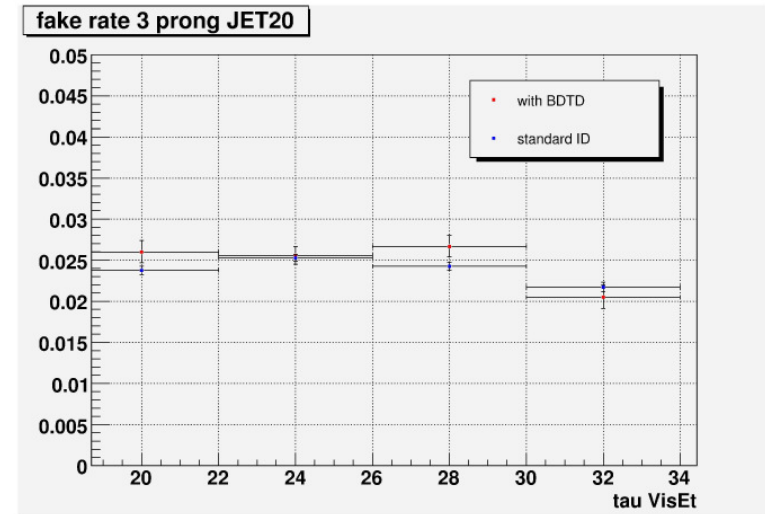
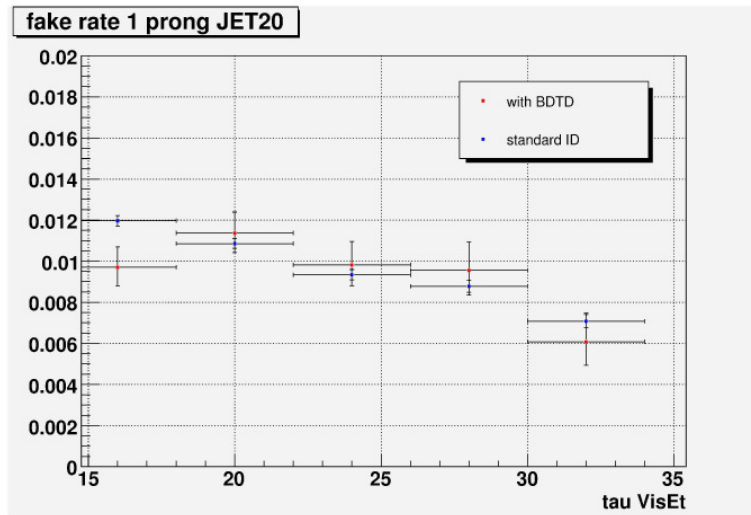


32 – 35 GeV  
# ev. 3200



**FINALLY...**  
**APPLICATION OF TRAINED BDT TO**  
**INDEPENDENT (EVENTS NOT USED FOR**  
**TRAINING) MC AND DATA SAMPLES**

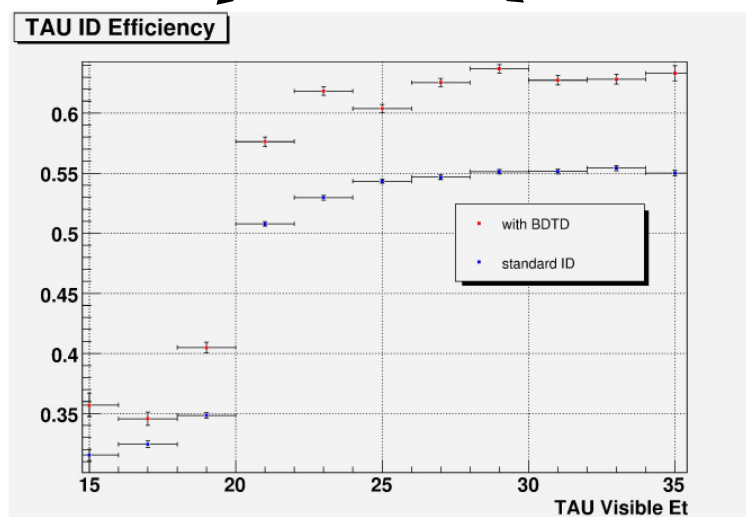
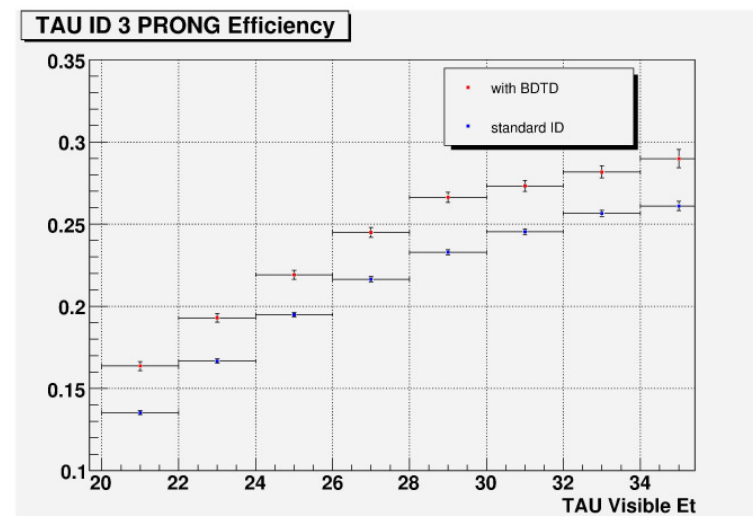
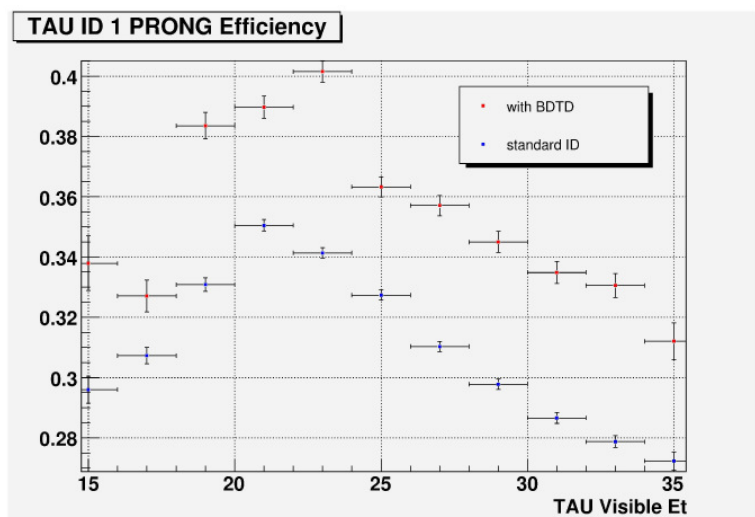
# FAKE RATE



Similar behaviour  
of fake rate, as desired

Note:  
these are fake rates  
relative to taufinder  
taus

# EFFICIENCY



Efficiency at plateau  
is increased of about  
15%!

Note:  
these are efficiencies  
relative to taufinder  
taus



## CONCLUSIONS

- **Results are really promising. We have under control jet fakes, and we increase tau efficiency of about 15%.**
- **The energy range considered up to now is limited (15 – 35 GeV), we plan to extend the study (with other QCD samples, e.g. JET\_50, JET\_70...)**
- **We plan to apply this algorithm to a  $Z \rightarrow \tau\tau$  selection, in order to check the real improvement of this new tau-ID algorithm on data**